

# Restoring STM images via Sparse Coding: noise and artifact removal

João P. Oliveira, Ana Bragança, José Bioucas-Dias, Mário Figueiredo, Luís Alcácer, Jorge Morgado, Quirina Ferreira

## Abstract

In this article, we present a denoising algorithm to improve the interpretation and quality of scanning tunneling microscopy (STM) images. Given the high level of self-similarity of STM images, we propose a denoising algorithm by reformulating the true estimation problem as a sparse regression, often termed sparse coding. We introduce modifications to the algorithm to cope with the existence of artifacts, mainly dropouts, which appear in a structured way as consecutive line segments on the scanning direction. The resulting algorithm treats the artifacts as missing data, and the estimated values outperform those algorithms that substitute the outliers by a local filtering. We provide code implementations for both Matlab and Gwyddion [19].

## I. INTRODUCTION

Scanning tunneling microscopy (STM) has been considered the most important technique to image, control and monitorize molecular systems with atomic resolution under a wide range of experimental conditions that enable molecule visualization. Typically, the STM image is composed of a bidimensional array of current values obtained when a metallic tip is very close to a conductive substrate creating a tunnel effect by applying a voltage between both conductive media or electrodes. The image resolution depends on several factors, such as the nature of electrodes, conductivity of the molecular system, packing order of molecules, and temperature. The quality of STM images is, however, limited by various degradation mechanisms including noise (thermal and Poissonian due to the electronic counting process), drifts, and dropouts. These degradation mechanisms reduce considerably the interpretability of STM images, calling for image denoising and restoration techniques to mitigate the effect of those degradations mechanisms.

There are several software suites that perform image denoise of STM images [19], [11]. Most techniques implement denoising in the two-dimensional frequency domain by exploiting the fact that the STM images are quasi-periodicity and thus have quasi-discrete and quasi-periodic patterns in the frequency domain.

However, the degradation mechanisms mentioned above defocus the frequency patterns, compromising the success of the frequency based denoising techniques.

Image denoising is one of the oldest image restoration problems, which continues to be object of intense research. As in many signal and image areas, the success of image denoising depends crucially on the existence of a good model to represent the original image. In this paper, we attack STM denoising by reformulating the true estimation problem as a sparse regression, often termed sparse coding. Following the standard procedure in patch-based image restoration, the image is partitioned into small overlapping square patches and the vector corresponding to each patch is modeled as a sparse linear combination of vectors, termed atoms, taken from a set called dictionary. Aiming at optimal sparse representations, and thus at optimal noise removing capabilities, the dictionary is learned from the data it represents via matrix factorization with sparsity constraints on the code (i.e., the regression coefficients) enforced by the  $\ell_1$  norm [6], [16].

The sparsity of a representation on a dictionary is strongly linked with the level of self-similarity of the images under study. It happens that STM images have a high level of self-similarity and, therefore, they are natural candidates to be denoised via sparse regression on learned dictionaries. Compared with the standard approaches, STM images pose an additional challenge: they have plenty of artifacts, mainly dropouts, which appear in a structured way as consecutive line segments on the scanning direction. To cope with these artifacts, we modify the algorithm introduced in [6], [16]; the modifications consist in treating the artifacts as missing data in both the dictionary learning and denoising (i.e., coding) phases.

To assess the quality of the proposed approach, we applied the algorithm to STM images of porphyrins and bipyridines adsorbed on highly oriented pyrolytic graphite. The Matlab code of the algorithm and a Gwyddion module can be obtained from <http://www.lx.it.pt/~jpaos/stm/>.

## II. SPARSE IMAGE REPRESENTATION

Signals and images can be represented in different ways. Several methods include Fourier Transform methods [24], wavelets methods [17], curvelets methods [23], just to mention a few. Despite the underlying differences, each representation has its own advantages when used in some of the classical, and still challenging, image restoration problems [12], [3].

Recently, sparse linear decomposition of signals in learned dictionaries led to state-of-the-art methods. The idea is to represent a signal, or fragments of the signal denoted patches, as a linear combination of vectors, termed atoms, taken from a set called dictionary. The dictionary can be fixed, for example based on wavelets [17], or can be learned [1].

The topic of sparse and redundant representations of real world images is currently one of the most active research fields. Recently, there has been an increased interest due to the fact that dictionaries yielding sparse representations may be learned directly from the data they represent [6], [16].

The fact that the real world images (and signals) admit sparse representations in suitable dictionaries is a consequence of the high level of self-similarity of real world images (and signals), i.e., given an image patch, there is a high likelihood of finding similar patches at different locations and scales. This is also true in the case of STM images: by nature, an STM image is almost a periodic image, and thus there exist similar patches with high probability.

The proposed algorithm works as follows: (i) the noisy STM image is decomposed in smaller patches; (ii) a dictionary is learned from the data; (iii) the patches are sparsely coding using just a few atoms from the learned dictionary; and (iv) the denoised/restored image is obtained by composition of the coded patches. In the following sections we introduce notation and describe in detail the constituent parts of the proposed *sparseland* framework to restore STM images.

#### A. Patch decomposition and composition of an image

Sensor measurements are usually corrupted by noise. STM images are not an exception, and the observation mechanism is well approximated by the additive model

$$\mathbf{z} = \mathbf{x} + \mathbf{n},$$

where  $\mathbf{z}$  is the noisy image,  $\mathbf{x}$  the true image, and  $\mathbf{n}$  models noise and model errors. The noise is originated by different sources, such as the tip movements, thermal noise, Poissonian counting noise, or even from the fact that different molecules have different conductivities. Given the relatively number of independent causes, we assume that the noise is zero-mean independent Gaussian with an unknown variance  $\sigma$ . The goal is to obtain an estimate of  $\mathbf{x}$  given  $\mathbf{z}$ .

Unlike other denoise algorithms, the proposed *sparseland* framework processes the image in small patches rather than the whole image. Consider the noisy image  $\mathbf{z} \in \mathbb{R}^N$ , where  $N = N_1 \times N_2$  is the image size, and the patch  $\mathbf{z}_i \in \mathbb{R}^m$  containing the pixels located inside a square window of size  $\sqrt{m} \times \sqrt{m}$  centered at the  $i$ -th pixel. The total number of overlapping patches that exist in image  $\mathbf{z}$  is  $N_p = (N_1 - \sqrt{m} + 1)(N_2 - \sqrt{m} + 1)$ . Let  $\mathbf{x}_i \in \mathbb{R}^m$  and  $\mathbf{n}_i \in \mathbb{R}^m$  denote two vectors holding, respectively, the elements  $\mathbf{x}$  and of  $\mathbf{n}$  corresponding to the  $i$ -th patch. We have then

$$\mathbf{z}_i = \mathbf{x}_i + \mathbf{n}_i, \quad i = 1, \dots, N_p.$$

In the *sparseland* framework, patches are processed individually, and thus for each  $\mathbf{x}_i$  we obtain an estimate  $\hat{\mathbf{x}}_i$ . Given that the reconstruction is not exact, we can write

$$\hat{\mathbf{x}}_i = \mathbf{x}_i + \varepsilon_i, \quad i = 1, \dots, N_p,$$

where  $\varepsilon_i$  is the estimation error for the  $i$ -th patch. To produce the overall estimate of  $\mathbf{x}$  from the estimates  $\mathbf{x}_i$ , for  $i = 1, \dots, N_p$ , we combine the information of several patches as follows. We introduce selection matrices  $\mathbf{S}_i$  such that  $\mathbf{x}_i = \mathbf{S}_i \mathbf{x}$ . Note that each row of  $\mathbf{S}_i$  contains just one non-null element of value 1. Consider also the following matrix and vectors:

$$\begin{aligned} \mathbf{S} &:= [\mathbf{S}_1^T, \dots, \mathbf{S}_{N_p}^T]^T \\ \hat{\mathbf{x}}_P &:= [\hat{\mathbf{x}}_1^T, \dots, \hat{\mathbf{x}}_{N_p}^T]^T \\ \varepsilon_P &:= [\varepsilon_1^T, \dots, \varepsilon_{N_p}^T]^T. \end{aligned}$$

With the above notation, the estimate of each patch is given by

$$\hat{\mathbf{x}}_P = \mathbf{S} \mathbf{x} + \varepsilon_P,$$

and the overall estimate  $\hat{\mathbf{x}}$  can be obtained by

$$\hat{\mathbf{x}} = \mathbf{S}^\dagger \hat{\mathbf{x}}_P, \tag{1}$$

where  $\mathbf{S}^\dagger := (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T$ . Noting that  $\mathbf{S}^T \mathbf{x}_i$  places the path number  $i$  at its position in the image and that  $(\mathbf{S}^T \mathbf{S})$  is a diagonal matrix whose  $i$ -th diagonal element holds the number of times pixel  $i$  appears in any patch, the estimate for the  $i$ -th pixel is the average of all its estimates, one per patch containing thereof. This condition implies that  $(\mathbf{S}^T \mathbf{S})$  is invertible as far as any pixel belongs at least to one patch.

The estimate given by (1) is the best linear unbiased estimator for  $\mathbf{x}$  [13], as long as we assume that  $\varepsilon_P$  is zero-mean and independent and identically distributed (iid), and thus having a covariance matrix proportional to the identity matrix, i.e.,  $\mathbf{C}_\varepsilon \approx \mathbf{I}$ . However, due to the overlapping structure of the patches it is very difficult to accurately compute  $\mathbf{C}_\varepsilon$ . We opted to use the suboptimal estimate (1) that, nevertheless, produces state-of-the-art results.

## B. Dictionary Learning

Given the periodic characteristic of a STM image, it is very likely that it admits a sparse representation in fixed existing dictionaries (i.e., representations). Wavelets and Fourier bases are two paradigmatic examples. However, the use of learned dictionaries from the data they represent holds systematically the

state-of-the-art [6], [15]. The goal of dictionary learning is to find a dictionary such that the patches can be accurately represented with just a few number of elements. One way to formulate this idea is solving the following optimization problem:

$$\min_{\mathbf{D} \in C, \alpha_1, \dots, \alpha_{N_p}} \sum_{i=1}^{N_p} (1/2) \|\mathbf{z}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1, \quad (2)$$

where  $\mathbf{D} \equiv [\mathbf{d}_1, \dots, \mathbf{d}_k] \in \mathbb{C}^{m \times k}$ ,  $C := \{\mathbf{D} \in \mathbb{C}^{m \times k} : |\mathbf{d}_j^T \mathbf{d}_j| \leq 1, j = 1, \dots, k\}$ . The quadratic terms account for the representation errors and the  $l_1$  norm is used to promote sparse codes. The relative weight between the two terms is established by the regularization parameter  $\lambda > 0$ . The constraint  $\mathbf{D} \in C$  is just a technical condition to prevent  $\mathbf{D}$  from being arbitrarily large.

Although the focus of this paper is not on the optimization of (2), it is worth mentioning just a few words about it. A common approach to handle these problems [14], [1], [20], [22] is to alternate the optimization with respect to  $\mathbf{D}$  and  $\alpha_1, \dots, \alpha_{N_p}$ . The optimization with respect to  $\mathbf{D}$  is a quadratic problem with convex constraints and optimization with respect to  $\alpha_1, \dots, \alpha_{N_p}$  is also convex and decoupled (we can solve individually for each  $\alpha_i$  variable). However, the optimization with respect to all variables is non-convex. Although we are not guaranteed to converge to the global minimum, the obtained stationary points have shown to produce state-of-the-art results.

In a typical STM image, we have  $N_p = 60000$ ,  $m = 100$ , and  $k = 64$ . These numbers mean that the computational burden necessary to compute (2) is heavy. To cope with this shortcoming, we adopt the online learning approach introduced in [15], which largely reduces that computational complexity. For more details see in [10].

### C. Sparse Coding

Once we have learned a dictionary  $\mathbf{D}$  with respect to which the patches of  $\mathbf{x}$  admit a sparse representation, we proceed and compute the estimate of a given patch  $\mathbf{x}_i = \mathbf{S}_i \mathbf{x}$ , for  $i \in \{1, \dots, N_p\}$ . Following the standard formulation in synthesis based approaches to sparse regression, the patch estimate  $\hat{\mathbf{x}}_i = \mathbf{D}\hat{\alpha}$  is the solution of the constrained optimization

$$\min_{\alpha} \|\alpha\|_0 \quad \text{subject to: } \|\mathbf{D}\alpha - \mathbf{z}_i\|^2 \leq \delta, \quad (3)$$

where  $\mathbf{z}_i = \mathbf{S}_i \mathbf{z}$  is the observed patch corresponding to the true patch  $\mathbf{x}_i$ ,  $\alpha \in \mathbb{C}^k$  and  $\|\alpha\|_0$  is the number of nonzero elements of vector  $\alpha$ , and  $\delta \geq 0$  is a parameter controlling the reconstruction error.

When  $\delta = 0$  and the dictionary is full column range, the solution of the optimization problem (3), if it exists, is easy to compute. However, in the majority of the situations, this is not the case because the

dictionaries are often overcomplete and, due to noise,  $\delta > 0$ . Under these circumstances, the problem is NP-hard [18]. Two different approaches have been followed. One consists in replacing  $\|\cdot\|_0$  by the  $l_1$  norm, which results in the so-called least absolute shrinkage and selection operator (LASSO) [25], which is equivalent to the basis pursuit denoising (BPDN) [4]. The other consists in attacking directly the original problem using greedy algorithms such as the orthogonal basis pursuit (OMP) [21], iterative hard thresholding (IHT) [2], hard thresholding pursuit [9], or approximate message passing (MP) [26].

In this work we follow the same approach as in [10], and choose the OMP [21] algorithm because of its lower computational complexity compared with the others. The details can be found in [10].

### III. NOISE AND ARTIFACT REMOVAL

Assuming that the noise of each patch is zero-mean and has covariance  $\sigma^2 \mathbf{I}$ , it is possible to show [10] that the relative attenuation of noise is given by

$$\frac{\mathbb{E}[\|\varepsilon_i\|_2^2]}{\mathbb{E}[\|n_i\|_2^2]} = \frac{p}{m}. \quad (4)$$

This means that the estimation error is proportional to the sparsity level of the signal. However, in practice, the ratio (4) is hardly achieved. One can also understand the noise reduction from a geometric point of view. The sparse coding of the patches, given by optimization problem (3), can be viewed as a projection of the noisy patch  $\mathbf{z}_i$  onto the subspace spanned by the active columns of  $\mathbf{D}$  in that patch, corresponding to the non-zero components of  $\alpha$ . Given that the number of active atoms is much smaller than the  $m$ , the noise is largely reduced.

The *sparseland* framework works well for STM images because the images are almost periodic. This makes it easier to learn the dictionary directly from the data. Besides the presence of noise, STM images usually have plenty of artifacts. However, they do not appear isolated, such as impulsive noise, but rather in a structured way. Typically, several consecutive pixels along a scanning line may be corrupted. Figure 1 shows an example of a STM image with artifacts that hampers the observation of darker spots. The image represents a polymorphic monolayer with zinc(II)octaethylporphyrin (ZnOEP) molecules organized as dimer-like units with different brightness aligned along parallel rows and absorbed on highly oriented pyrolytic graphite (HOPG) by  $\pi - \pi$  stacking interaction [8]. The difference in brightness of the two ZnOEP units is due to their different coupling to the HOPG where the darker molecules have lower signal and for this reason is hard to distinguish with high detail the central part of porphyrines and their axial groups.

In order to proceed, special care must be taken to eliminate these artifacts, otherwise the learned dictionary ends up with patches that can represent well these outliers. To overcome these problems, we

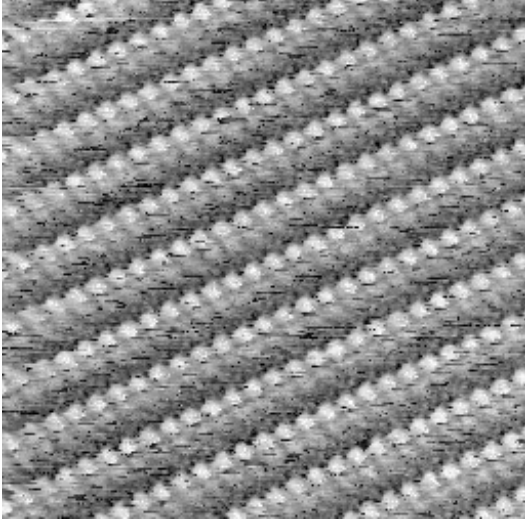


Fig. 1. STM image ( $30 \times 30$  nm,  $V=570$  mV,  $I=28$  pA) of a polymorphic phase of a ZnOEP monolayer absorbed on HOPG where the ZnOEP are organized as dimers with different brightness. The image has important artifacts that make it hard to interpret.

change the algorithm by introducing an extra vector that masks out these pixels. This approach has several advantages. It makes it easy to iteratively improve on current results, either by adding or removing pixels from the mask (considered outliers). It also allows us to easily fill in the outliers pixels (inpainting) with new values. As opposed to a classical local filtering method, the new inpainting values can take into account the full image statistics, and thus attain superior reconstruction quality.

#### A. Proposed algorithm

The proposed algorithm changes the previously introduced *sparseland* framework to accomplish two objectives: (i) deal with the structured outliers; and (ii) do inpainting of those pixels. We start by defining a binary vector  $\mathbf{m} \in \{0, 1\}^N$  of size  $N = N_1 \times N_2$ , where 0 indicates that the corresponding pixel in  $\mathbf{x}$  is considered an outlier. The purpose of this vector is to mask out the outliers. Let  $\mathbf{m}_i = \mathbf{S}_i \mathbf{m}$  for  $i \in \{1, \dots, N_p\}$  denote a patch of  $\mathbf{m}$ . Note that  $\mathbf{m}_i$  indicate the outlier's positions in the patches  $\mathbf{x}_i$ . Figure 2 shows the STM image of Figure 1 with the mask over-imposed. The pixels considered outliers are represented in red.

The dictionary learning algorithm is the same as described in section II-B. To avoid learning the outliers structures, instead of considering all the patches, we use  $\mathbf{m}_i$  to discard those that contain outliers. The sparse coding algorithm is also the same as described in section II-C. However, in this case, we do not

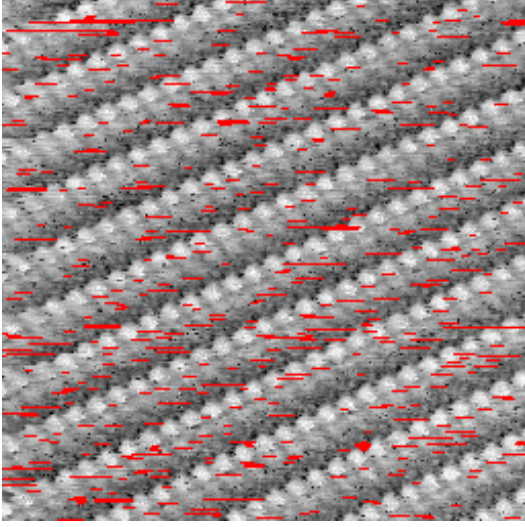


Fig. 2. STM image of Figure 1 with masked outliers represented in red.

discard any patch; we code the patches ignoring only the pixels that are considered outliers by solving the optimization

$$\min_{\alpha} \|\alpha\|_0 \quad \text{subject to: } \|\mathbf{m}_i \odot (\mathbf{D}\alpha - \mathbf{z}_i)\|^2 \leq \delta, \quad (5)$$

where the symbol  $\odot$  represents elementwise multiplication. Note that the effect of  $\mathbf{m}_i \odot \mathbf{D}$  and  $\mathbf{m}_i \odot \mathbf{z}_i$  is to project the columns of  $\mathbf{D}$  and  $\mathbf{z}_i$  onto a lower dimension subspace. The corresponding change in the OMP algorithm is simply an element-wise multiplication by  $\mathbf{m}_i$ .

Finally, once we have a sparse representation for all the patches, the image composition is the pixel-wise average of all the patches covering a pixel, marked or not as an outlier. The inpainting values obtained outperform those that result from substituting the outliers by a local filtering. In section IV we compare both approaches. A Matlab version and a Gwyddion[19] module of the code are available at <http://www.lx.it.pt/~jpaos/stm/>.

### B. Pre-processing and artifact selection

Before we apply the proposed algorithm, raw STM images are pre-processed and artifacts are identified, i.e., matrix  $\mathbf{m}$  is defined. We start by applying the following pre-processing steps:

- 1) mean level subtraction: a plane is adjusted to all the data points and subtracted from the data;
- 2) outlier removal: pixel data that is below 1% or above 99% of the quantile are removed;
- 3) leveling by median subtraction: lines are adjusted to match height median.



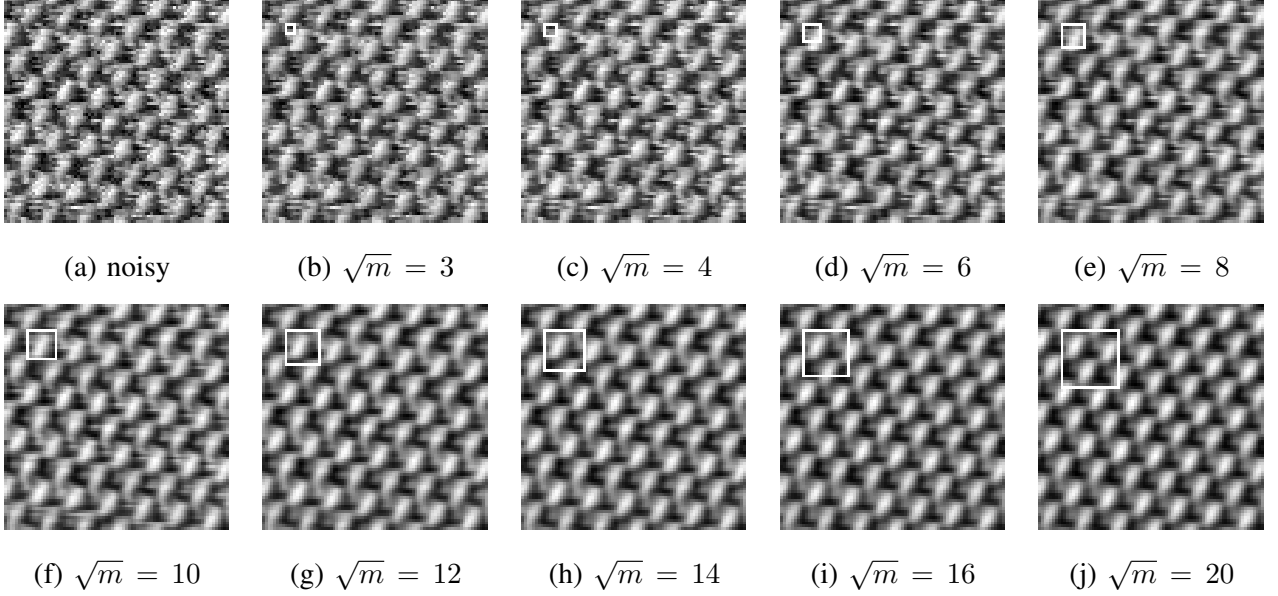


Fig. 3. (a) STM image ( $10 \times 10 \text{ nm}$ ,  $V=0.56 \text{ V}$ ,  $I=0.2 \text{ nA}$ ) of an incomplete BP monolayer absorbed on a full packed monolayer of ZnOEP that is absorbed on HOPG substrate. Each bright region correspond to a molecular wire formed by HOPG/ZnOEP/BP alternated by unreacted ZnOEP molecules (darker domains); (b)–(j) close up of STM image denoise with different patch sizes. The small white square represents the patch size at the image’s scale.

Following this pre-processing steps, artifacts are identified by inspection and marked in matrix  $\mathbf{m}$ . Remark that common image processing software suites have tools designed to obtain  $\mathbf{m}$  automatically. As an example, Gwyddion [19] implements an algorithm denoted *Mark Scars*, where several parameters for outlier detection can be specified such as minimum/maximum length and width.

For better results, the process can be iterated: start with an initial mask, apply the algorithm, improve on the mask from restored image and repeat the process.

### C. Parameter selection

There are several parameters that affect the behavior of the overall algorithm. In this section we discuss how these parameters can be tuned in order to obtain the best possible outcome. The algorithm is made up of two parts: dictionary learning and sparse coding.

The first parameter that we need to define is the patch size, that is common for both parts. Patch size influences the amount of noise that is possible to remove. Bigger patch sizes results in better noise reduction. However, the patch size influences how data can be represented by the dictionary. Smaller patch sizes means that the dictionary adapts better to data, including details and noise, while larger patch sizes capture the similarities and periodic components.

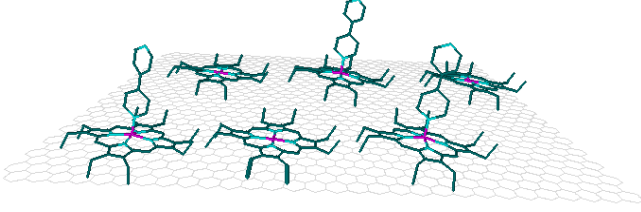


Fig. 4. Molecular schema illustrating the HOPG/ZnOEP/BP molecular wires alternated by ZnOEP absorbed on HOPG.

From the exhaustive experiments, we concluded that the patch size should be at least equal to the smallest interest object that can be represented by  $\sqrt{m} \times \sqrt{m}$  pixels. Figure 3 (a) shows an incomplete monolayer with 4,4'-bipyridine (BP) absorbed on top of a full packed monolayer of ZnOEP through a Zn-N bond. Each pair of ZnOEP on BP forms a vertical molecular wire absorbed on a HOPG substrate [7]. The brighter regions are attributed to the BP molecules bonded to Zn alternated with darker domains that correspond to unreacted ZnOEP molecules. The corresponding molecular schema is depicted in Figure 4.

Figures 3 (b)–(k) show the results obtained by the proposed algorithm by varying the patch size. As it can be seen, smaller patch sizes adapt better to the noise image, and thus attain smaller noise reduction. On the other hand, larger patch sizes result in an almost periodic image, where important details are lost. For this particular image — Figure 3 —, the best results are obtained when the patch size is around  $\sqrt{m} = 10$ . In all experiments the regularization parameter was set at  $\lambda = 0.11$ , and the dictionary consisted of 64 atoms.

For the sparse coding algorithm (5), two parameters need to be set: the maximum number of atoms (MNA) and the error parameter  $\delta$ . These parameters determine the reconstruction *error* of patch  $\mathbf{z}_i$ : although 1 or 2 patches aren't enough to correctly represent data, a larger number gives more flexibility and adaptability, and thus less noise reduction. The implementation of algorithm [10] searches for the sparsest number of atoms that attain an error less than  $\delta$ , with at most the MNA. Note that the MNA is a hard threshold, i.e., each patch is not reconstructed with more atoms than that.

The error parameter  $\delta$  depends on the noise statistics of the image. Based on a Gaussian noise distribution assumption, an estimate [10] for  $\delta$  is given by  $\delta = \sigma^2 F_{\chi^2(m)}^{-1}(\gamma)$ , where the noise variance  $\sigma$  is estimated by the MAD rule [5]. In all the experiments we set  $\gamma = 0.96$ .

#### IV. RESULTS

In this section we assess the merit of the proposed algorithm applied on the STM image of polymorphic phase of a ZnOEP monolayer absorbed on HOPG, previously depicted in Figure 1. We selected for the

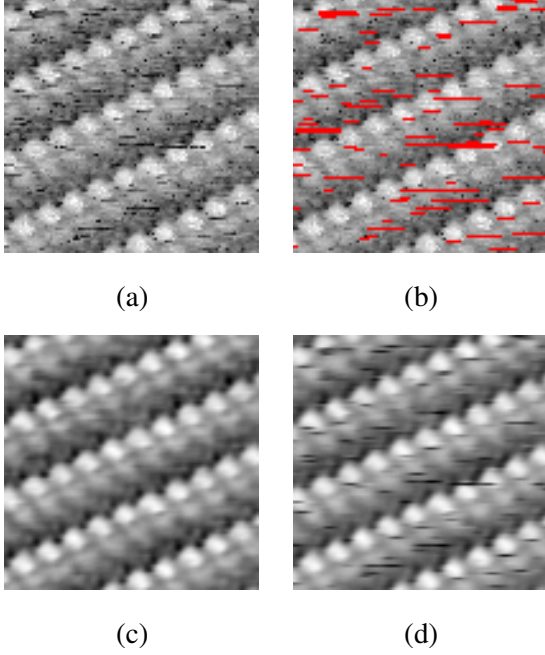


Fig. 5. Close up of STM image ( $30 \times 30 \text{ nm}$ ,  $V=570 \text{ mV}$ ,  $I=28 \text{ pA}$ ) of a polymorphic phase of a ZnOEP monolayer absorbed on HOPG where the ZnOEP are organized as dimers with different brightness: (a) noisy image; (b) noisy image with mask (in red); (c) restored with proposed algorithm; (d) restored with proposed algorithm without mask.

patch size  $\sqrt{m} = 10$ , the maximum number of atoms was set at 4 and  $\delta = 10$ . The mask used is depicted in Figure 2.

To compare the inpainting results, we also run a version of the proposed algorithm without the mask. Close ups of the noisy image, mask and restored images are depicted in Figure 5.

In Figure 5 (c), the restored image (denoised) shows clearly the details of darker ZnOEP where it is possible to observe the ZnOEP central part surrounded by 4 protrusions that we attribute to the ethyl groups that are oriented towards the HOPG as is represented by molecular schema of Figure 6.

When the proposed algorithm is used without a mask, Figure 5 (d), the structured outliers are only partially removed. This happens because the dictionary learning algorithm actually learns these structures. Thus, there will be atoms available in  $\mathbf{D}$  such that a patch with outliers still admits a sparse representation. However, when a mask is used, the algorithm removes these outliers and fill in the pixels with new values. Instead of a local median or a more sophisticated interpolation algorithm, by relying on the learned dictionary, the filled in values are in accordance to the global image statistics.

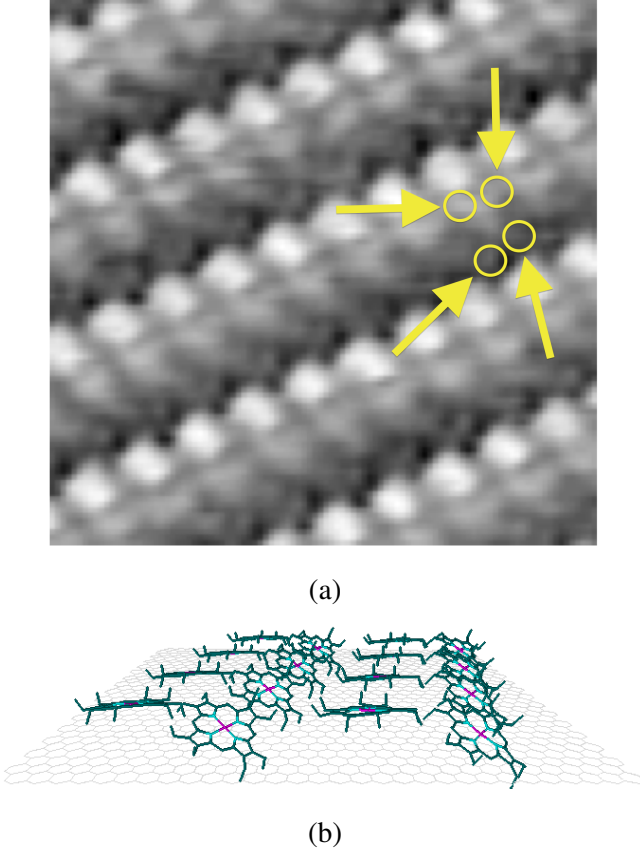


Fig. 6. (a) Detail of ZnOEP central part surrounded by 4 protrusions; and (b) monolayer molecular schema where the ZnOEP molecules that are parallel to the HOPG correspond to the brighter domains and the tilted ones correspond to the darker domains.

## V. CONCLUSION

We presented a denoising algorithm, formulated as sparse regression problem, to mitigate the effect of the degradation mechanisms observed in STM images. The proposed formulation was introduced based on the fact that STM images present a high level of self-similarity, so that a suitable dictionary could be learned from the noisy image. The algorithm also cope with the existence of artifacts, mainly dropouts, by treating them as missing data. The results obtained outperform those algorithms that substitute the outliers by a local filtering, as the filled in values are now in accordance to the global image statistics.

## ACKNOWLEDGEMENT

The authors thank FCT-Portugal for financial support under the project PEst-OE/EEI/LA0008/2013 and Post-Doc grant to QF.

## REFERENCES

- [1] M. Aharon, M. Elad, and A. Bruckstein, “K-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [2] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” *Appl. Comp. Harm. Anal.*
- [3] A. Bovik, Ed., *Handbook of Image & Video Processing*. Academic Press, 2000.
- [4] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [5] D. L. Donoho and J. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [6] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.
- [7] Q. Ferreira, A. Bragança, L. Alcácer, and J. Morgado, “Conductance of well-defined porphyrin self-assembled molecular wires up to 14 nm in length,” *Journal of Physical Chemistry C*, vol. 118, no. 3, pp. 7229–7234, March 2014.
- [8] Q. Ferreira, A. Bragança, O. N. O. Oliveira, N. M. Moura, M. Faustino, L. Alcácer, and J. Morgado, “Dynamics of porphyrin adsorption on highly oriented pyrolytic graphite monitored by scanning tunnelling microscopy at the liquid/solid interface,” *Applied Surface Science*, vol. 273, no. n.a., pp. 220–225, May 2013.
- [9] S. Foucart, “Hard thresholding pursuit: an algorithm for compressive sensing,” *SIAM Journal on Numerical Analysis*, vol. 49, no. 6, pp. 2543–2563, 2011.
- [10] H. Hongxing, J. M. Bioucas-Dias, and V. Katkovnik, “Interferometric phase estimation via sparse coding in the complex domain,” *IEEE Transactions on Geoscience and Remote Sensing*, Submitted 2013.
- [11] I. Horcas, R. Fernández, J. M. Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, and A. M. Baro, “Wsxm: A software for scanning probe microscopy and a tool for nanotechnology,” *Review of Scientific Instruments*, vol. 78, no. 1, pp. –, 2007. [Online]. Available: <http://scitation.aip.org/content/aip/journal/rsi/78/1/10.1063/1.2432410>
- [12] A. Jain, *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.
- [13] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993, vol. 1.
- [14] H. Lee, A. Battle, R. Raina, and A. Y. Ng, “Efficient sparse coding algorithms,” in *In NIPS*. NIPS, 2007, pp. 801–808.
- [15] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding,” in *International Conference on Machine Learning*, 2009.
- [16] J. Mairal, M. Elad, and G. Sapiro, “Sparse representation for color image restoration,” *Image Processing, IEEE Transactions on*, vol. 17, no. 1, pp. 53–69, Jan 2008.
- [17] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press Inc, 1999.
- [18] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995. [Online]. Available: <http://dx.doi.org/10.1137/S0097539792240406>
- [19] D. Nečas and P. Klapetek, “Gwyddion: an open-source software for SPM data analysis,” *Central European Journal of Physics*, vol. 10, pp. 181–188, 2012.
- [20] B. A. Olshausen and D. J. Fieldt, “Sparse coding with an overcomplete basis set: a strategy employed by v1,” *Vision Research*, vol. 37, pp. 3311–3325, 1997.
- [21] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 1–3.

- [22] A. Rakotomamonjy, “Applying alternating direction method of multipliers for constrained dictionary learning,” *Neurocomputing*, vol. 106, no. 0, pp. 126 – 136, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231212008612>
- [23] J.-L. Starck, E. Candes, and D. Donoho, “The curvelet transform for image denoising,” *Image Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 670–684, Jun 2002.
- [24] E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, 1971.
- [25] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.
- [26] J. Vila and P. Schniter, “Expectation-maximization Bernoulli-Gaussian approximate message passing,” in *in Proc. Asilomar Conf. Signals Syst. Comput*, 2011.